BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37164
doi: 10.3897/biss.3.37164

OPEN ACCESS

Conference Abstract

Use of European Open Science Cloud and National e-Infrastructures for the Long-Term Storage of Digitised Assets from Natural History Collections

Abraham Nieva de la Hidalga[‡], Nicolas Cazenave[§], Donat Agosti[|], Zhengzhe Wu[¶], Mathias Dillen[#], Lars H Nielsen[¤]

‡ Cardiff University School of Computer Science and Informatics, Cardiff, United Kingdom
§ CINES, Montpellier, France
| www.plazi.org, Bern, Switzerland
¶ University of Helsinki, Helsinki, Finland
# Meise Botanic Garden, Meise, Belgium
¤ CERN/Zenodo, Geneve, Switzerland

Corresponding author: Abraham Nieva de la Hidalga (nievadelahidalgaa@cardiff.ac.uk)

Received: 12 Jun 2019 | Published: 18 Jun 2019

Citation: Nieva de la Hidalga A, Cazenave N, Agosti D, Wu Z, Dillen M, Nielsen L (2019) Use of European Open Science Cloud and National e-Infrastructures for the Long-Term Storage of Digitised Assets from Natural History Collections. Biodiversity Information Science and Standards 3: e37164. https://doi.org/10.3897/biss.3.37164

Abstract

Digitisation of Natural History Collections (NHC) has evolved from transcription of specimen catalogues in databases to web portals providing access to data, digital images, and 3D models of specimens. These portals increase global accessibility to specimens and help preserve the physical specimens by reducing their handling. The size of the NHC requires developing high-throughput digitisation workflows, as well as research into novel acquisition systems, image standardisation, curation, preservation, and publishing. Nowadays, herbarium sheet digitisation workflows (and fast digitisation stations) can digitise up to 6,000 specimens per day. Operating those digitisation stations in parallel, can increase the digitisation capacity. The high-resolution images obtained from these specimens, and their volume require substantial bandwidth, and disk space and tapes for storage of original digitised materials, as well as availability of computational processing resources for generating derivatives, information extraction, and publishing. While large institutions have dedicated digitisation teams that manage the whole workflow from

© Nieva de la Hidalga A et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

acquisition to publishing, other institutions cannot dedicate resources to support all digitisation activities, in particular long-term storage. National and European e-infrastructures can provide an alternative solution by supporting different parts of the digitisation workflows. In the context of the Innovation and consolidation for large scale digitisation of natural heritage (ICEDIG Project 2018), three different e-infrastructures providing long-term storage have been analysed through three pilot studies: EUDAT-CINES, Zenodo, and National Infrastructures.

The EUDAT-CINES pilot centred on transferring large digitised herbarium collections from the National Museum of Natural History France (MNHN) to the storage infrastructure provided by the Centre Informatique National de l'Enseignement Supérieur (CINES 2014), a European trusted digital repository. The upload, processing, and access services are supported by a combination of services provided by the European Collaborative Data Infrastructure (EUDAT CDI 2019) and CINES . The Zenodo pilot included the upload of herbarium collections from Meise Botanic Garden (APM) and other European herbaria into the Zenodo repository (Zenodo 2019). The upload, processing and access services are supported by Zenodo services, accessed by APM. The National Infrastructures pilot facilitated the upload of digital assets derived from specimens of herbarium and entomology collections held at the Finnish Museum of Natural History (LUOMUS) into the Finnish Biodiversity Information Facility (FinBIF 2019). This pilot concentrates on simplifying the integration of digitisation facilities to Finnish national e-infrastructures, using services developed by LUOMUS to access FinBIF resources.

The data models employed in the pilots allow defining data schemas according to the types of collection and specimen images stored. For EUDAT-CINES, data were composed of the specimen data and its business metadata (those the institution making the deposit, in this case MNHN, considers relevant for the data objects being stored), enhanced by archiving metadata, added during the archiving process (institution, licensing, identifiers, project, archiving date, etc). EUDAT uses ePIC identifiers (ePIC 2019) to identify each deposit. The Zenodo pilot was designed to allow defining specimen data and metadata supporting indexing and access to resources. Zenodo uses DataCite Digital Object Identifiers (DOI) and the underlying data types as the main identifiers for the resources, augmented with fields based on standard TDWG vocabularies. FinBIF compiles Finnish biodiversity information to one single service for open access sharing. In FinBIF, HTTP URI based identifiers are used for all data, which link the specimen data with other information, such as images.

The pilot infrastructure design reports describe features, capacities, functions and costs for each model, in three specific contexts are relevant for the implementation of the Distributed Systems of Scientific Collections (DiSSCo 2019) research infrastructure, informing the options for long-term storage and archiving digitised specimen data. The explored options allow preservation of assets and support easy access. In a wider context, the results provide a template for service evaluation in the European Open Science Cloud (EOSC 2019) which can guide similar efforts.

Keywords

European Open Science Cloud, e-infrastructure, long-term storage, federated cloud, digitisation, natural history collections, ICEDIG, EUDAT, Zenodo

Presenting author

Abraham Nieva de la Hidalga

Presented at

Biodiversity_Next 2019

Funding program

Horizon 2020 Framework Programme of the European Union

Grant title

ICEDIG – “Innovation and consolidation for large scale digitisation of natural heritage” H2020-INFRADEV-2016-2017 – Grant Agreement No. 777483

Hosting institution

EUDAT, CINES, Zenodo, Plazi, Botanic Gardens Meise, University of Finland

References

- CINES (2014) Centre Informatique National de l’Enseignement Supérieur (National Computer Center for Higher Education) . https://www.cines.fr/. Accessed on: 2019-5-01.
- DiSSCo (2019) DiSSCo - Distributed System of Scientific Collections Research Infrastructure. https://dissco.eu/. Accessed on: 2019-5-01.
- EOSC (2019) European Open Science Cloud Portal. https://www.eosc-portal.eu/. Accessed on: 2019-5-01.
- ePIC (2019) ePIC Persistent Identifiers for eResearch. https://www.pidconsortium.eu/. Accessed on: 2019-5-01.
- EUDAT CDI (2019) EUDAT Research Data Services, Expertise & Technology Solutions. https://eudat.eu/. Accessed on: 2019-5-01.

- FinBIF (2019) Finnish Biodiversity Information Facility. https://laji.fi/en. Accessed on: 2019-5-01.
- ICEDIG Project (2018) Innovation and consolidation for large scale digitisation of natural heritage Grant agreement ID:777483. H2020 programme. Start: 2018-01-01 End: 2020-03-31. https://cordis.europa.eu/project/rcn/216630/factsheet/en. Accessed on: 2019-5-01.
- Zenodo (2019) Zenodo - Research Shared. https://zenodo.org/. Accessed on: 2019-5-01.